An alternatif test
to the two independent samples t test

Rohmatul Fajriyah

Institute of Statistics, TU Graz, Austria

Dept. of Statistics, Univ Islam Indonesia, Jogjakarta, Indonesia

fajriyah@student.tugraz.at, rfajriyah@fmipa.uii.ac.id

Abstract

In this paper we proposed the alternative test to the two independent and normally distributed samples t test based on the cross variance concept. We present the simulation results of the power and the error rate of the special case of the cross variance which is when the variances of the two samples are homogeneous and the t tests.

The simulation results show that the special case of the cross variance test has the power and the error type I rate equal to the t test. This result suggests that the proposed test could be used as an alternative to detect whether there is difference between means of the two independent normally distributed samples. We give some example comparative case studies of the special case of the cross variance and the t tests.

Key Words: cross variance; error rate type I; power test; t test.

1 Introduction

When we have two independent samples and would like to detect whether there is difference between means of the two populations then we apply the parametric and the non parametric tests. If the samples come from normal distribution, then the t test will be the right choice.

In this paper we will introduce a cross variance concept and present a new approach to detect whether there are differences between the means of the two samples based on the cross variance.

This paper is organized as follows. In Section 2 we will describe the cross-variance concept and the proposed test, inlcuding the new probability density function. Section 3 will present the simulation results of the power and the error type I of the proposed test in comparison to the t test. Some examples will be provided together with the t test. Finally, section 4 is the summary and remarks.

2 An introduction to the cross variance concept and the proposed test

2.1 The cross variance concept

Definition 2.1. *Suppose we have two independent samples, X_i and Y_j; $i = 1, 2, ..., m$ and $j = 1, 2, ..., n$. Their sample mean and variance are denoted by $\overline{X}, \overline{Y}$ and V_x, V_y. Let*

$$V_x^* = \frac{\sum_{i=1}^{m}(X_i - \overline{Y})^2}{m-1} \quad and \quad V_y^* = \frac{\sum_{i=1}^{n}(Y_i - \overline{X})^2}{n-1}$$

be the cross-variance for each sample X and Y respectively. The cross-variance sample of groups $\boldsymbol{X}$ and $\boldsymbol{Y}$ is defined as

$$\boldsymbol{T} = \frac{V_x^a + V_y^a}{2}, \tag{1}$$

where $\quad V_x^a = \frac{V_x}{V_x^*}, \quad V_y^a = \frac{V_y}{V_y^*}.$

Clearly

$$\begin{aligned} V_x^* &= \frac{\sum_{i=1}^{m}(X_i - \overline{Y})^2}{m-1} \\ &= V_x + \frac{m(\overline{Y} - \overline{X})^2}{m-1}, \end{aligned} \tag{2}$$

and

$$\begin{aligned} V_y^* &= \frac{\sum_{i=1}^{n}(Y_i - \overline{X})^2}{n-1} \\ &= V_y + \frac{n(\overline{Y} - \overline{X})^2}{n-1}. \end{aligned}$$

Thus $T = \frac{V_x^a + V_y^a}{2}$ can be re-written as

$$\begin{aligned} T &= \frac{1}{2}\left[\frac{V_x}{V_x + \frac{m(\overline{Y}-\overline{X})^2}{m-1}} + \frac{V_y}{V_y + \frac{n(\overline{Y}-\overline{X})^2}{n-1}}\right] \\ &= \frac{V_x}{2V_x + 2\frac{m(\overline{Y}-\overline{X})^2}{m-1}} + \frac{V_y}{2V_y + 2\frac{n(\overline{Y}-\overline{X})^2}{n-1}} \\ &= Z_1 + Z_2 \end{aligned} \tag{3}$$

In what follows, we assume that

1. the sample sizes are equal

2. X_i and Y_i are i.i.d. normally distributed with unknown means and known variances σ_x^2, σ_y^2.

It follows that

$$\frac{(n-1)V_x}{\sigma_x^2} \sim \chi_{(n-1)}^2, \ \frac{(n-1)V_y}{\sigma_y^2} \sim \chi_{(n-1)}^2, \ \text{and} \ \frac{n(\overline{Y}-\overline{X})^2}{\sigma_y^2 + \sigma_x^2} \sim \chi_{(1)}^2$$

Therefore Equation (3) can be written as follows

$$T = Z_1 + Z_2 \tag{4}$$

$$= \frac{\frac{(n-1)V_x}{\sigma_x^2}}{2\frac{(n-1)V_x}{\sigma_x^2} + 2\frac{(\sigma_x^2+\sigma_y^2)}{\sigma_x^2}\frac{n}{(\sigma_x^2+\sigma_y^2)}(\overline{Y}-\overline{X})^2} + \frac{\frac{(n-1)V_y}{\sigma_y^2}}{2\frac{(n-1)V_y}{\sigma_y^2} + 2\frac{(\sigma_x^2+\sigma_y^2)}{\sigma_y^2}\frac{n}{(\sigma_x^2+\sigma_y^2)}(\overline{Y}-\overline{X})^2}$$

where

$$Z_1 = \frac{\frac{(n-1)V_x}{\sigma_x^2}}{\frac{2(n-1)V_x}{\sigma_x^2} + \frac{2(\sigma_x^2+\sigma_y^2)}{\sigma_x^2}\frac{n}{(\sigma_x^2+\sigma_y^2)}(\overline{Y}-\overline{X})^2} = \frac{U}{2U + 2abV},$$

and

$$Z_2 = \frac{\frac{(n-1)V_y}{\sigma_y^2}}{\frac{2(n-1)V_y}{\sigma_y^2} + \frac{2(\sigma_x^2+\sigma_y^2)}{\sigma_y^2}\frac{n}{(\sigma_x^2+\sigma_y^2)}(\overline{Y}-\overline{X})^2} = \frac{S}{2S + 2bcV}$$

with

$$U = \frac{(n-1)V_x}{\sigma_x^2}, \quad S = \frac{(n-1)V_y}{\sigma_y^2}, \quad V = \frac{n(\overline{Y}-\overline{X})^2}{(\sigma_x^2 + \sigma_y^2)}$$

and

$$a = \frac{1}{\sigma_x^2}, \quad b = (\sigma_x^2 + \sigma_y^2), \quad c = \frac{1}{\sigma_y^2}.$$

Hence Equation (4) can be written as

$$T = Z_1 + Z_2 = \frac{U}{2U + 2abV} + \frac{S}{2S + 2bcV} \tag{5}$$

To compute the distribution of T in Equation (5), it can be done by considering the fact that

1. U, V and S are independent

3

2. Z_1 and Z_2 are dependent

In this paper we will describe the first case, where we consider that U, V and S are independent.

2.2 The proposed test

Under normality asumption of X and Y then U, V and S are independent, where V is $\chi^2_{(1)}$ distributed and U, S are $\chi^2_{(n-1)}$ distributed. From the equation (5), suppose $V = Z_3$ from here we get that $U = \frac{2abZ_1Z_3}{1-2Z_1}$ and $S = \frac{2bcZ_2Z_3}{1-2Z_2}$. The jacobian of this transformation is

$$
|J| = \begin{vmatrix} \frac{dU}{dZ_1} & \frac{dU}{dZ_2} & \frac{dU}{dZ_3} \\ \frac{dS}{dZ_1} & \frac{dS}{dZ_2} & \frac{dS}{dZ_3} \\ \frac{dV}{dZ_1} & \frac{dV}{dZ_2} & \frac{dV}{dZ_3} \end{vmatrix} = \begin{vmatrix} \frac{2abZ_3}{(1-2Z_1)^2} & 0 & \frac{2abZ_1}{(1-2Z_1)} \\ 0 & \frac{2bcZ_3}{(1-2Z_2)^2} & \frac{2bcZ_2}{(1-2Z_2)} \\ 0 & 0 & 1 \end{vmatrix} = \frac{4ab^2cZ_3^2}{\left((1-2Z_1)(1-2Z_2)\right)^2}
$$

The joint probability density function of Z_1, Z_2, Z_3 is

$$
\begin{aligned}
&f_{Z_1,Z_2,Z_3}(z_1, z_2, z_3) \\
&= f_U\left(u = \frac{2abz_1z_3}{1-2z_1}\right) f_S\left(s = \frac{2bcz_2z_3}{1-2z_2}\right) f_V(v = z_3)|J| \\
&= \frac{\left(4ab^2c\right)^{\frac{n-1}{2}}}{2^{n-\frac{1}{2}}\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)^2} \frac{z_1^{\frac{n-1}{2}-1} z_2^{\frac{n-1}{2}-1} z_3^{\frac{2n-1}{2}-1} e^{-z_3\left(\frac{1}{2}+\frac{abz_1}{1-2z_1}+\frac{cbz_2}{1-2z_2}\right)}}{(1-2z_1)^{\frac{n+1}{2}} (1-2z_2)^{\frac{n+1}{2}}}
\end{aligned}
$$

$$(6)$$

The joint density function of Z_1, Z_2 is the marginal probability function of Z_1, Z_2

from the equation (6) above. It is computed as follows:

$$f_{Z_1,Z_2}(z_1, z_2)$$

$$= \int_0^\infty f_{Z_1,Z_2,Z_3}(z_1, z_2, z_3)dz_3$$

$$= \frac{\left(4ab^2c\right)^{\frac{n-1}{2}}}{2^{n-\frac{1}{2}}\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)^2} \frac{z_1^{\frac{n-1}{2}-1} z_2^{\frac{n-1}{2}-1}}{(1-2z_1)^{\frac{n+1}{2}}(1-2z_2)^{\frac{n+1}{2}}} \int_0^\infty z_3^{\frac{2n-1}{2}-1} e^{-z_3\left(\frac{1}{2}+\frac{abz_1}{1-2z_1}+\frac{bcz_2}{1-2z_2}\right)} dz_3$$

$$= \frac{\left(4ab^2c\right)^{\frac{n-1}{2}}\Gamma\left(n-\frac{1}{2}\right) z_1^{\frac{n-1}{2}-1} z_2^{\frac{n-1}{2}-1}}{2^{n-\frac{1}{2}}\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)^2 (1-2z_1)^{\frac{n+1}{2}}(1-2z_2)^{\frac{n+1}{2}}\left(\frac{1}{2}+\frac{abz_1}{(1-2z_1)}+\frac{bcz_2}{(1-2z_2)}\right)^{n-\frac{1}{2}}}$$

$$(7)$$

Therefore the cumulative distribution function (cdf) of $T \le t$ s computed as follows

$$F_T(t)$$

$$= P(T \le t)$$

$$= \int_{-\infty}^\infty \int_{-\infty}^{t-z_1} f_{Z_1,Z_2}(z_1, z_2)dz_2 dz_1$$

$$= \int_0^t \int_0^{t-z_1} \frac{\left(4ab^2c\right)^{\frac{n-1}{2}}\Gamma\left(n-\frac{1}{2}\right) z_1^{\frac{n-1}{2}-1} z_2^{\frac{n-1}{2}-1}dz_2 dz_1}{2^{n-\frac{1}{2}}\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)^2 \left((1-2z_1)(1-2z_2)\right)^{\frac{n+1}{2}}\left(\frac{1}{2}+\frac{abz_1}{(1-2z_1)}+\frac{bcz_2}{(1-2z_2)}\right)^{n-\frac{1}{2}}}$$

$$= \int_0^t \frac{\left(4ab^2c\right)^{\frac{n-1}{2}}\Gamma\left(n-\frac{1}{2}\right) z_1^{\frac{n-1}{2}-1}}{2^{n-\frac{1}{2}}\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)^2 (1-2z_1)^{\frac{n+1}{2}}}\mathbf{B}dz_1 \qquad (8)$$

where

$$\mathbf{B} = \int_0^{t-z_1} \frac{z_2^{\frac{n-1}{2}-1}(1-2z_2)^{\frac{n+1}{2}} dz_2}{\left(\frac{1}{2}+\frac{abz_1}{(1-2z_1)}+\frac{bcz_2}{(1-2z_2)}\right)^{n-\frac{1}{2}}} \qquad (9)$$

To compute the integral at Equation (9), first we will simplify this $\left(\frac{1}{2}+\frac{abz_1}{(1-2z_1)}+\frac{bcz_2}{(1-2z_2)}\right)^{n-\frac{1}{2}}$ as $\left(1+\frac{2(bc-(1-2(1-ab-bc)z_1))}{1-2(1-ab)z_1}z_2\right)\left(\frac{1-2(1-ab)z_1}{2(1-2z_1)}\right)$. Therefore the

Equation (9) becomes

$$\frac{\left(2(1-2z_1)\right)^{n-\frac{1}{2}}}{\left(1-2(1-ab)z_1\right)^{n-\frac{1}{2}}} \int\limits_{0}^{t-z_1} \frac{z_2^{\frac{n-1}{2}-1}\left(1-2z_2\right)^{\frac{n+1}{2}} dz_2}{\left(1+\frac{2(bc-(1-2(1-ab-bc)z_1))}{1-2(1-ab)z_1}z_2\right)^{n-\frac{1}{2}}} \tag{10}$$

The integral $\int\limits_{0}^{t-z_1} \frac{z_2^{\frac{n-1}{2}-1}(1-2z_2)^{\frac{n+1}{2}} dz_2}{\left(1+\frac{2(bc-(1-2(1-ab-bc)z_1))}{1-2(1-ab)z_1}z_2\right)^{n-\frac{1}{2}}}$ is written as

$$\int\limits_{0}^{t-z_1} z_2^{\frac{n-1}{2}-1}\left(1-2z_2\right)^{\frac{n+1}{2}}\left(1+\frac{2\big(bc-(1-2(1-ab-bc)z_1)\big)}{1-2(1-ab)z_1}z_2\right)^{-\left(n-\frac{1}{2}\right)} dz_2 \tag{11}$$

By considering the binomial expansion then Equation (11) can be represented as

$$=\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\left[\frac{2^{k+l}(-1)^{k+l}\binom{\frac{n+1}{2}}{k}\binom{n-\frac{3}{2}+l}{l}(bc-1+2(1-ab-bc)z_1)^l}{\left(1-2(1-ab)z_1\right)^l}\times\right.$$
$$\left.\int\limits_{0}^{t-z_1} z_2^{\frac{n-1}{2}+k+l} dz_2\right]$$
$$=\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\left[\frac{2^{k+l}(-1)^{k+l}\binom{\frac{n+1}{2}}{k}\binom{n-\frac{3}{2}+l}{l}(bc-1+2(1-ab-bc)z_1)^l}{\left(1-2(1-ab)z_1\right)^l}\times\right.$$
$$\left.\frac{(t-z_1)^{\frac{n+1}{2}+k+l}}{\frac{n+1}{2}+k+l}\right] \tag{12}$$

Therefore,

$$\mathbf{B} = \frac{\left(2(1-2z_1)\right)^{n-\frac{1}{2}}}{\left(1-2(1-ab)z_1\right)^{n-\frac{1}{2}}} \times$$

$$\left[\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\left[\frac{2^{k+l}(-1)^{k+l}\binom{\frac{n+1}{2}}{k}\binom{n-\frac{3}{2}+l}{l}(bc-1+2(1-ab-bc)z_1)^l}{\left(1-2(1-ab)z_1\right)^l} \times\right.\right.$$

$$\left.\left.\frac{(t-z_1)^{\frac{n+1}{2}+k+l}}{\frac{n+1}{2}+k+l}\right]\right] \tag{13}$$

and

$$F_T(t) = \frac{\left(4ab^2c\right)^{\frac{n-1}{2}}\Gamma\left(n-\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)^2} \times$$

$$\left[\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\frac{(-2)^k\binom{\frac{n+1}{2}}{k}\binom{n-\frac{3}{2}+l}{l}(-2(bc-1))^l t^{\frac{n+1}{2}+k+l}}{\frac{n+1}{2}+k+l}\mathbf{G}\right] \tag{14}$$

where

$$\mathbf{G} = \int_0^t\left[z_1^{\frac{n-1}{2}-1}\left(1-\frac{z_1}{t}\right)^{\frac{n+1}{2}+k+l}(1-2z_1)^{\frac{n}{2}-1}\left(1-2(1-ab)z_1\right)^{-(n+l-\frac{1}{2})}\times\right.$$

$$\left.\left(1+\frac{2(1-ab-bc)z_1}{bc-1}\right)^l dz_1\right] \tag{15}$$

Again, by considering the binomial expansion then Equation (15) can be written as

follows

$$
\mathbf{G} = \sum_{m=0}^{\infty} \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} \left[\binom{\frac{n}{2}}{m} \binom{l}{p} \binom{n+l+q-\frac{3}{2}}{q} (-2)^m \left(\frac{2(1-ab-bc)}{bc-1} \right)^p (-2(1-ab))^q \times \right.
$$

$$
\left. \int_0^t z_1^{\frac{n-1}{2+m+p+q}-1} \left(1 - \frac{z_1}{t} \right)^{\frac{n-1}{2}+k+l} dz_1 \right]
$$

$$
= \sum_{m=0}^{\infty} \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} \left[\binom{\frac{n}{2}}{m} \binom{l}{p} \binom{n+l+q-\frac{3}{2}}{q} (-2)^m \left(\frac{2(1-ab-bc)}{bc-1} \right)^p (-2(1-ab))^q \times \right.
$$

$$
\left. \frac{B\left(\frac{n-1}{2} + m + p + q, \frac{n+1}{2} + k + l \right)}{t^{\frac{n-1}{2}+m+p+q}} \right] \tag{16}
$$

Therefore $F_T(t)$ is

$$
F_T(t) = \frac{\left(4ab^2c \right)^{\frac{n-1}{2}} \Gamma\left(n - \frac{1}{2} \right)}{\Gamma\left(\frac{1}{2} \right) \Gamma\left(\frac{n-1}{2} \right)^2} \quad \times
$$

$$
\left[\sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} \left[\frac{2^k \binom{\frac{n-1}{2}+k}{k} \binom{n-\frac{3}{2}+l}{l} (-2(bc-1))^l t^{\frac{n-1}{2}+k+l}}{\frac{n-1}{2} + k + l} \quad \times \right. \right.
$$

$$
\frac{\binom{\frac{n}{2}}{m} \binom{l}{p} \binom{n+l+q-\frac{3}{2}}{q} (-2)^m \left(\frac{2(1-ab-bc)}{bc-1} \right)^p (-2(1-ab))^q}{t^{\frac{n-1}{2}+m+p+q}} \quad \times
$$

$$
\left. \left. B\left(\frac{n-1}{2} + m + p + q, \frac{n+1}{2} + k + l \right) \right] \right]
$$

$$
= \frac{\left(4ab^2c \right)^{\frac{n-1}{2}} \Gamma\left(n - \frac{1}{2} \right)}{\Gamma\left(\frac{1}{2} \right) \Gamma\left(\frac{n-1}{2} \right)^2} \quad \times
$$

$$
\left[\sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} \left[\frac{2^k \binom{\frac{n-1}{2}+k}{k} \binom{n-\frac{3}{2}+l}{l} (-2(bc-1))^l t^{k+l-m-p-q}}{\frac{n-1}{2} + k + l} \quad \times \right. \right.
$$

$$
\binom{\frac{n}{2}}{m} \binom{l}{p} \binom{n+l+q-\frac{3}{2}}{q} (-2)^m \left(\frac{2(1-ab-bc)}{bc-1} \right)^p (-2(1-ab))^q \quad \times
$$

$$
\left. \left. B\left(\frac{n-1}{2} + m + p + q, \frac{n+1}{2} + k + l \right) \right] \right] \tag{17}
$$

Furthermore, from the Equation (17) then the probability density function (pdf) of

T is

$$f_T(t) = \frac{\left(4ab^2c\right)^{\frac{n-1}{2}}\Gamma\left(n-\frac{1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n-1}{2}\right)^2}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{m=0}^{\infty}\sum_{p=0}^{\infty}\sum_{q=0}^{\infty}\frac{2^k\binom{\frac{n-1}{2}+k}{k}\binom{n-\frac{3}{2}+l}{l}}{\frac{n-1}{2}+k+l}\quad\times$$

$$(k+l-m-p-q)t^{k+l-m-p-q}(-2(bc-1))^l\binom{\frac{n}{2}}{m}\binom{l}{p}\binom{n+l+q-\frac{3}{2}}{q}(-2)^m\times$$

$$\left(\frac{2(1-ab-bc)}{bc-1}\right)^p(-2(1-ab))^q B\left(\frac{n-1}{2}+m+p+q,\frac{n+1}{2}+k+l\right)\quad(18)$$

The pdf of T also can be computed as follows:

$$f_T(t) = \int_{-\infty}^{t} f_{Z_1,Z_2}(z_1, t-z_1)dz_1$$

$$= \int_{-\infty}^{t} f_{Z_1,Z_2}(t-z_2, z_2)dz_2 \qquad (19)$$

Now we already have the pdf and cdf of T from where we can compute the T statistics value for the hypothesis testing, that is the hypothesis null about the equality of mean of two-groups independent samples is rejected if $T \le t_0$ or $P(T \le t_0) = F_T(t_0) \le \alpha$.

The computation of $F_T(t_0)$ by using the Equation (18) is involving the five summation and hence it is not quite simple. The computation gets simple in the case $\sigma_x^2 = \sigma_y^2$, which will be described next.

2.3 Special case of the proposed test

In the case of $\sigma_x^2 = \sigma_y^2$, we estimate the V_x and V_y by the least square estimator of the pooled variance $S_p^2 = \frac{V_x+V_y}{2}$.

If we use the least square estimate as the estimator of V_x and V_y therefore the Equa-

tions (3) and (5) becomes

$$T^* = \frac{\frac{V_x+V_y}{2}}{\left[\frac{V_x+V_y}{2} + \frac{n\left(\overline{Y}-\overline{X}\right)^2}{n-1}\right]} \tag{20a}$$

$$= \frac{U^*}{U^* + 4V^*} \tag{20b}$$

where $U^* = \frac{(n-1)\left(V_x+V_y\right)}{\sigma_x^2}$ and $V^* = \frac{n\left(\overline{Y}-\overline{X}\right)^2}{2\sigma_x^2}$.

The pdf of T^* is derived from the ratio of linear combination of chi-square random variables [1]. First, let $Y = 1+4\frac{V^*}{U^*}$, where V^* is distributed $\chi^2_{(1)}$ and U^* is distributed $\chi^2_{2(n-1)}$. Second, the pdf of T^* is computed by taking $T^* = \frac{1}{Y}$.

In the folowing computation, the chi-square distribution will be represented as the Gamma distribution. Therefore, we have that V^* is Gamma distributed with parameters $\alpha_1 = \frac{1}{2}$ and $\beta_1 = 2$. U^* is Gamma distributed with parameters $\alpha_2 = (n-1)$ and $\beta_2 = 2$. U^* and V^* are independents.

Suppose $G = \frac{V^*}{U^*}$ and if we take $U^* = H$, then we get $V^* = GH$. Further we got the Jacobian of this transformation variable random is h. Because V^* and U^* are independents, then the joint probability function of G and H is

$$f_{G,H}(g,h) = f_{V^*,U^*}(gh,h).h \tag{21}$$

where

$$f_{V^*,U^*}(gh,h) = f_{V^*}(gh).f_{U^*}(h),$$

$$f_{V^*}(gh) = \frac{(gh)^{\alpha_1-1}e^{-\frac{gh}{\beta_1}}}{\beta_1^{\alpha_1}\Gamma(\alpha_1)}, \text{and}$$

$$f_{U^*}(h) = \frac{(h)^{\alpha_2-1}e^{-\frac{h}{\beta_2}}}{\beta_2^{\alpha_2}\Gamma(\alpha_2)}$$

10

Therefore

$$f_{G,H}(g,h) = \frac{(gh)^{\alpha_1-1}e^{-\frac{gh}{\beta_1}}}{\beta_1^{\alpha_1}\Gamma(\alpha_1)} \cdot \frac{(h)^{\alpha_2-1}e^{-\frac{h}{\beta_2}}}{\beta_2^{\alpha_2}\Gamma(\alpha_2)}h$$

$$= \frac{g^{\alpha_1-1}h^{\alpha_1+\alpha_2-1}e^{(-\frac{(1+g)}{\beta}h)}}{\beta^{\alpha_1+\alpha_2}\Gamma(\alpha_1)\Gamma(\alpha_2)}, \quad \beta_1 = \beta_2 = \beta \quad (22)$$

We want to determine the pdf of g then

$$f_G(g) = \int_0^\infty f_{G,H}(g,h)dh$$

$$= \frac{g^{\alpha_1-1}}{B(\alpha_1,\alpha_2)(1+g)^{\alpha_1+\alpha_2}}$$

$$= \frac{g^{\frac{1}{2}-1}}{B\left(\frac{1}{2},n-1\right)(1+g)^{n-\frac{1}{2}}} \quad (23)$$

G is distributed beta of second kind.

The next step is determining the distribution of Y. We define $Y = 1 + 4G$ and by using the transformation random variable, where $G = \frac{Y-1}{4}$, the pdf of Y is computed as follow

$$f_Y = \frac{\left(\frac{y-1}{4}\right)^{\frac{1}{2}-1}}{B\left(\frac{1}{2},n-1\right)\left(1+\left(\frac{y-1}{4}\right)\right)^{n-\frac{1}{2}}}\frac{1}{4}$$

$$= \frac{4^{n-1}(y-1)^{\frac{1}{2}-1}}{B\left(\frac{1}{2},n-1\right)(y+3)^{n-\frac{1}{2}}}, \quad 1 \le y \le \infty \quad (24)$$

The pdf of $T^* = \frac{1}{Y}$ is obtained

$$f_{T^*}(t^*) = \frac{4^{n-1}(\frac{1}{t^*}-1)^{\frac{1}{2}-1}}{B\left(\frac{1}{2},n-1\right)(3+\frac{1}{t^*})^{n-\frac{1}{2}}}\frac{1}{t^{*2}}$$

$$= \frac{4^{n-1}t^{n-2}(1-t^*)^{\frac{1}{2}-1}}{B\left(\frac{1}{2},n-1\right)(1+3t^*)^{n-\frac{1}{2}}}, \quad 0 \le t^* \le 1 \quad (25)$$

Furthermore the cdf of T^* analtically is computed as

$$
\begin{aligned}
F_{T^*}(t_0^*) &= \int_0^{t_0^*} f_{T^*}(t^*)dt^* \\
&= \frac{3^{n-1}}{B\left(\frac{1}{2}, n-1\right)} \int_0^{t_0^*} \frac{t^{*(n-2)}(1-t)^{\frac{1}{2}-1}}{(1+3t^*)^{n-\frac{1}{2}}} dt^* \\
&= \frac{4^{n-1}}{B\left(\frac{1}{2}, n-1\right)} \sum_{k=0}^{\infty} (-3)^k \binom{n-\frac{1}{2}+k-1}{k} \int_0^{t_0^*} t^{*(n-1+k-1)}(1-t^*)^{\frac{1}{2}-1} dt^* \\
&= \frac{4^{n-1}}{B\left(\frac{1}{2}, n-1\right)} \left[\sum_{k=0}^{\infty} (-3)^k \binom{n+k-\frac{3}{2}}{k} B\left(n-1+k, \frac{1}{2}\right) \times \right. \\
&\qquad \left. \text{pbeta}\left(t_0^*, n-1+k, \frac{1}{2}\right) \right]
\end{aligned}
\tag{26}
$$

We reject the null hypothesis of the equality of mean of two independent samples if $T^* < t_{0,\alpha}^*$ or $P\left(t^* < T_0^*\right) = F_{T^*}(T_0^*) = \text{p-value} \le P\left(t^* < t_{0,\alpha}^*\right) = F_{T^*}(t_{0,\alpha}^*) = \alpha$.

Observing that $(2(n-1))\left(\frac{V^*}{U^*}\right)$ is the square of a random variable having $t_{2(n-1)}$ distribution, a simple calculation shows that the same holds for the random variable

$$
J = \sqrt{(n-1)\left(\frac{1}{T^*} - 1\right)}
\tag{27}
$$

This statistic J can also be used to test the hypothesis $\mu_x = \mu_y$ and the critical values can be computed from the t table. It also follows that the J has a limiting normal distribution as $n \to \infty$

3   Simulation study

In this section we will describe the results from the two simulation studies of the t and the proposed tests. First simulation is done to measure the power of the proposed test is conducted at $N = 500$ times and $\alpha = 0.01$. The results are presented at Section

3.1. In this simulation, we consider various possibilities regarding the sample sizes and variances. The results of the simulation are divided into groups according to

1. variance: low ($S = 0.2$), medium ($S = 1.2$) and high ($S = 7$),

2. sample size: low (5 and 25), medium (100) and high (500)

Section 3.2 describes the second simulation results on measuring the rejection's rate under the null hypothesis of the t and the proposed tests, based on the $N = 500$ times simulation and $\alpha = 0.01$. In the simulation, we choose $\mu_X = \mu_Y = \mu = 9.2$ and $\sigma_X = \sigma_Y = \sigma$ are three values of variance $1.25, 3.5$ and 10 which represents the low, medium and high variance respectively.

Section 3.3 gives some examples on how the proposed test works on the data.

3.1   Power of the test

The simulation study to compute the power of the proposed and the t tests is conducted as follows:

1. Choose the $\mu_X, \mu_Y, \sigma_X = \sigma_Y = \sigma$ of the two-groups independent samples

2. The simulation under null hypothesis

    (a) Generate n random samples normally distributed with mean and standard deviation, μ_{X_0} and σ

    (b) Generate n random samples normally distributed with mean and standard deviation, μ_{Y_0} and σ

    (c) Compute their mean and variance samples

    (d) Compute t_0^* values, based on the equation (20a).

3. The simulation under alternative hypothesis

    (a) Generate n random samples normally distributed with mean and standard deviation, μ_{X_1} and σ

    (b) Generate n random samples normally distributed with mean and standard deviation, μ_{Y_1} and σ

    (c) Compute their mean and variance samples

    (d) Compute t_1^* values, based on the equation (20a).

4. Repeat steps (2) - (3), M times

5. Compute $t_{0,\alpha}^*$, the α quantile of t_0^*. Note that $t_{0,\alpha}^*$ also can be computed by using the αth quantile of pdf of T^* in the Equation (21).

6. Compute the power test of the proposed method $= 1 - \frac{\text{sum}(t_1^* \geq t_{0,\alpha}^*)}{M}$

7. Compute the power of t test from samples X_1 and Y_1

8. Compare the results from steps (6) and (7)

9. Do steps (1)-(8), for different value of mean and variance

Figures 3.1, 3.2, 3.3 and 3.4 show the power of the special case of the cross variance and t tests, based on the simulation study. They show that the proposed and the t tests have an equal power. In the computation, the power of the proposed test is provided by implementing the the empirical approach, therefore the results is not as smooth as the t test.
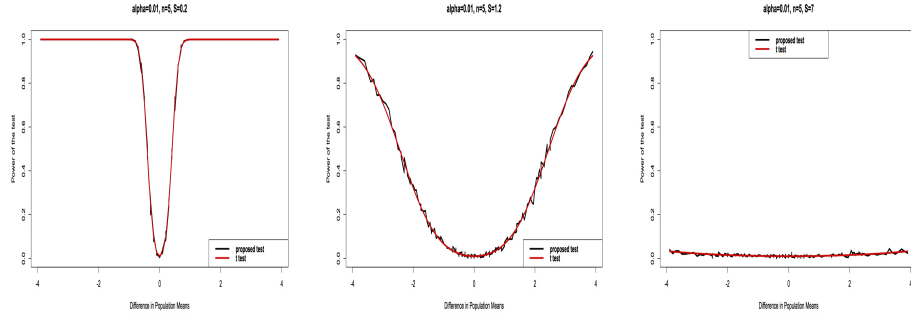
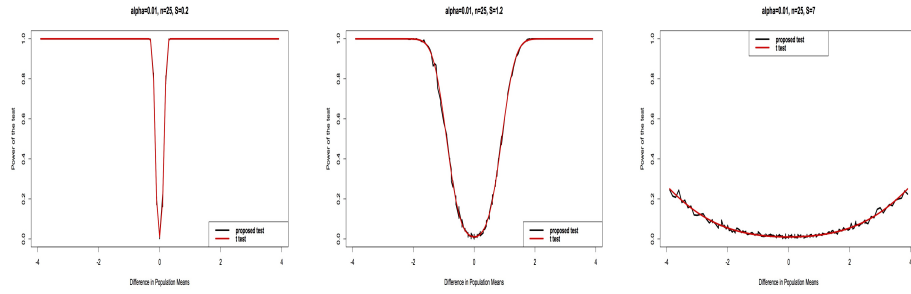Figure 3.1: Graphical power both the t and proposed tests, n=5

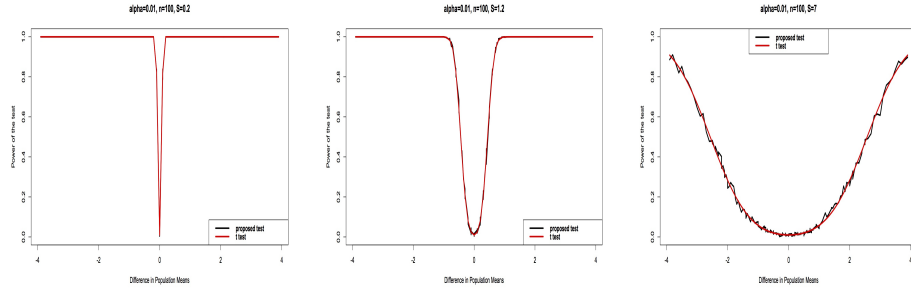Figure 3.2: Graphical power both the t and proposed tests, n=25

Figure 3.3: Graphical power both the t and proposed tests, n=100

3.2   Error type I rate

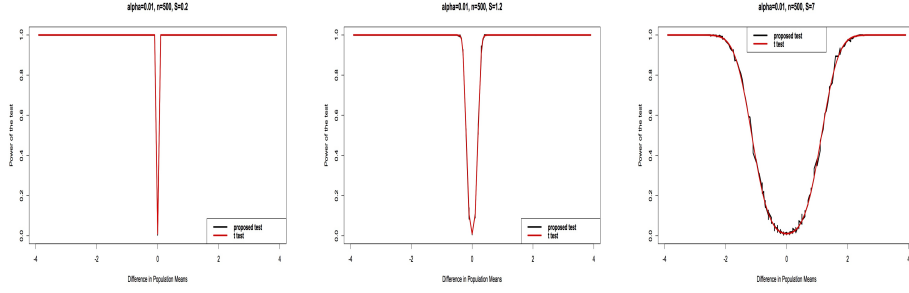The simulation study to compute the error type I rate is conducted as follows:

Figure 3.4: Graphical power both the t and proposed tests, n=500

1. Choose the $n, M_1, \alpha, \mu_X = \mu_Y = \mu$ and $\sigma_X = \sigma_Y = \sigma$ of the two-groups independent samples

2. Repeat M_1 times

    (a) Compute p-value of t_o^* test

    (b) Compute p-value of t test

3. Compute the proportion of t test p-value $< \alpha$ in M_1 results

4. Compute the proportion of the proposed test p-value $< \alpha$ in M_1 results

5. Compare the result of steps 3 and 4

In Table 1, it is shown the results of the simulation where the error type I rate of the proposed and the t tests are equal.

The distribution of p-values from the proposed and t tests are can be seen in the Figures 3.5, 3.6 and 3.7.

The next section will give some examples of data analysis concerning the proposed test and compare the results with the t test. We used 14 artificial data sets which the first 10 were taken randomly from the internet. The data can be seen in Table 2.

Table 1: Error type I rate under 500 simulation for the proposed and t tests

Sample size	Variance	proposed test		t test	
		0.05	0.01	0.05	0.01
	low	0.056	0.012	0.056	0.012
5	medium	0.062	0.016	0.062	0.016
	high	0.056	0.020	0.056	0.020
	low	0.046	0.010	0.046	0.010
25	medium	0.044	0.012	0.044	0.012
	high	0.038	0.010	0.038	0.010
	low	0.058	0.006	0.058	0.006
100	medium	0.050	0.010	0.050	0.010
	high	0.062	0.012	0.062	0.012
	low	0.038	0.004	0.038	0.004
500	medium	0.038	0.002	0.038	0.002
	high	0.050	0.010	0.050	0.010

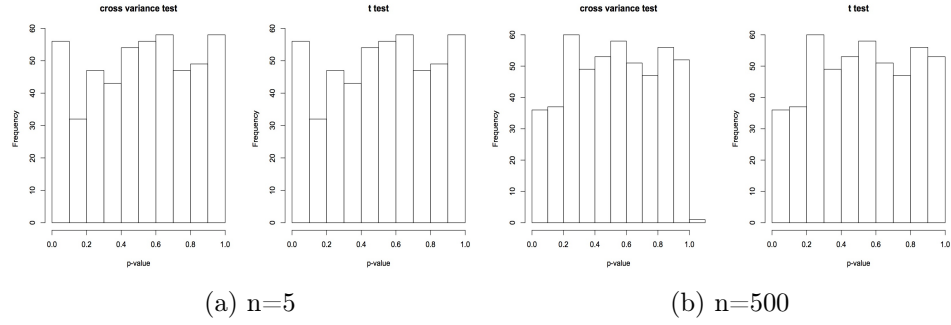(a) n=5   (b) n=500

Figure 3.5: P-values distribution of the proposed and t tests, small variance
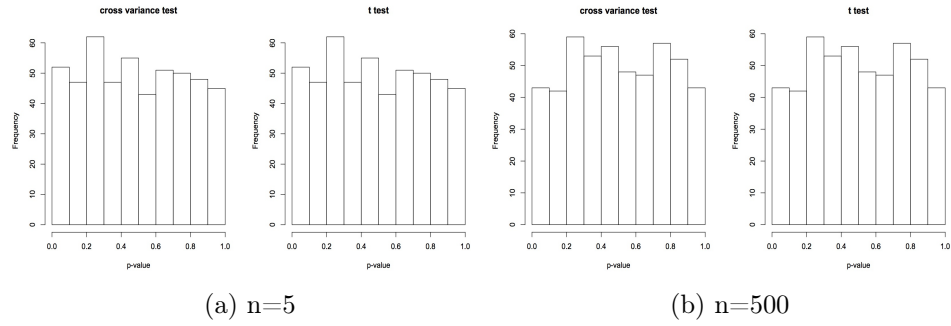

(a) n=5   (b) n=500

Figure 3.6: P-values distribution of the proposed and t tests, medium variance

3.3   Some examples

In this section, some examples data sets are provided and used as an example how to take the decision by using the special case of the cross variance test T^* ($=$ the T test
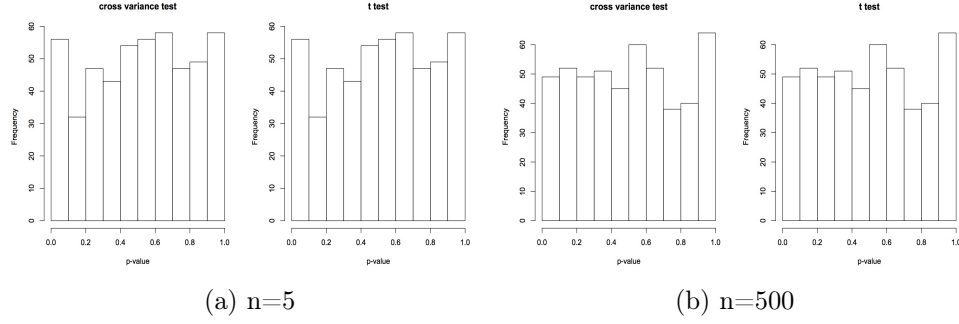
(a) n=5   (b) n=500

Figure 3.7: P-values distribution of the proposed and t tests, high variance

when $\sigma_X = \sigma_Y$). For this T^* test, the first step is making sure the variances of the samples are equal. There are some tests for this purpose, in this paper we implement the F.test in R.

The data sets are shown as in Table 2 and the computation results for the F.test are shown in Table 3. Tables 4 and 5 provide the results of the decision for both methods: the proposed and the t tests.

In the proposed test, the null hypothesis is rejected if $t_o^* < t_\alpha^*$ or if the p-value of the special case of the cross variance test less than α, where t_o^* is the statistics from the observed sample and p-value $= P\left(t_o^* < t_\alpha^*\right)$. In the computation we use $\alpha = 0.01$.

From Table 3 we can infer that data set 13 has unequal variances, therefore it will be excluded from further computation.

From the Table 4 we can see that the p-values and decisions from the proposed and the t tests are equal, except for the data set 4. It is the only one data set which has different sample size. When the sample size of two samples is different, we provide two options of n: the $\max(n_1, n_2)$ or the average between (n_1, n_2). The example of this is shown in Table 5.

4 Summary and Remarks

In this paper we have introduced the cross-variance concept, a new test based on the cross variance, the special case of the cross variance (as the proposed test) and the new probability density functions. A simulation study has been conducted to compute the power and the error type 1 rate of the proposed and the t tests.

The simulation study of the power shows that the proposed and the t tests has the same power. Furthermore the p-value and the error type I rate under the null hypothesis of the proposed and t tests are exactly equal. These results suggest that the proposed test could be used as an alternative test to detect whether there are difference between the means of two independent normal populations, where the variance and sample size is equal. In case the sample size is unequal we proposed to choose $n = \max(n_1, n_2)$ or the average of (n_1, n_2).

Conflict of interest: -

References

References

[1] Serge Provost and Edmund M. Rudiuk. The exact density function of the ratio of two dependent linear combinations of chi-square variables. *Annals of the Institute of Statistical Mathematics*, 46(3):557–571, 1994.

Table 2: Data sets, their mean and variance

Data	Data	Mean	Variance
1	X=(5,7,5,3,5,3,3,9)	5.000	4.571
	Y=(8,1,4,6,6,4,1,2)	4.000	6.571
2	X=(0.72,0.68,0.69,0.66,0.57,0.66,0.70,0.63,0.71,0.73)	0.675	0.002
	Y=(0.71,0.83,0.89,0.57,0.68,0.74,0.75,0.67,0.80,0.78)	0.742	0.008
3	X=(42,45,40,37,41,41,48,50,45,46)	43.500	15.833
	Y=(43,51,56,40,32,54,51,55,50,48)	48.000	57.330
4	X=(33,31,34,38,32,28)	32.667	11.067
	Y=(35,42,43,41)	40.250	12.917
5	X=(35,40,12,15,21,14,46,10,28,48,16,30,32,48,31,22, 12,39,19,25)	27.150	156.450
	Y=(2,27,38,31,1,19,1,34,3,1,2,3,2,1,2,1,3,29,37,2)	11.950	213.524
6	X=(26 21,22,26,19,22,26,25,24,21,23,23,18,29,22)	23.133	8.552
	Y=(18, 23, 21,20,20,29,20,16,20,26,21,25,17,18,19)	20.867	12.552
7	X=(520,460,500,470)	487.500	758.333
	Y=(230,270,250,280)	257.500	491.667
8	X=(3,0,6,7,4,3,2,1,4)	3.333	5.000
	Y=(5,1,5,7,10,9,7,11,8)	7.000	9.250
9	X=(16,20,21,22,23,22,27,25,27,28)	23.100	13.878
	Y=(19,22,24,24,25,25,26,26,28,32)	25.100	11.878
10	X=(91,87,99,77,88,91)	88.833	51.367
	Y=(101,110,103,93,99,104)	101.667	31.867
11	X=(10.11,7.36,6.34,11.83,8.61)	8.850	4.761
	Y=(3.28, 6.52,2.28,6.66,4.55)	4.658	3.760
12	X=(4.79,4.95,2.52,4.98,4.99)	4.446	1.166
	Y=(7.90,7.51,6.62,7.57,7.49)	7.418	0.227
13	X=(3.99,3.98,4.03,4.06, 3.84)	3.980	0.007
	Y=(6.68,6.25,6.97,5.75,4.01))	5.932	1.366
14	X=(10.16,8.26,16.23,1.44,0.66)	7.350	41.816
	Y=c(28.06,8.52,25.39,15.45,16.03)	18.690	63.422

Table 3: F.test decision

Samples	Decision	Samples	Decision
Data 1	equal variance	Data 8	equal variance
Data 2	equal variance	Data 9	equal variance
Data 3	equal variance	Data 10	equal variance
Data 4	equal variance	Data 11	equal variance
Data 5	equal variance	Data 12	equal variance
Data 6	equal variance	Data 13	not equal variance
Data 7	equal variance	Data 14	equal variance

Table 4: P-values and decision from the proposed and t tests

Sample	t test		proposed test	
	p-value	Decision	p-value	Decision
Data 1	0.411	Accept	0.411	Accept
Data 2	0.054	Accept	0.054	Accept
Data 3	0.114	Accept	0.113	Accept
Data 4	0.009	Reject		
Data 5	0.001	Reject	0.001	Reject
Data 6	0.067	Accept	0.066	Accept
Data 7	0.000	Reject	0.000	Reject
Data 8	0.010	Accept	0.010	Accept
Data 9	0.229	Accept	0.229	Accept
Data 10	0.006	Reject	0.006	Reject
Data 11	0.012	Accept	0.012	Accept
Data 12	0.000	Reject	0.000	Reject
Data 14	0.039	Accept	0.039	Accept

Table 5: P-values and decision from the proposed test, Data set 4

Choice of n	Least Square	
	p-value	Decision
Min	0.021	Accept
Max	0.004	Reject
Average	0.009	Reject